Classical justification of the Hall effect

Hamid V. Ansari
Department of Physics, Isfahan University, Isfahan, IRAN
Personal address: No. 16, Salman-Farsi Lane, Zeinabieh Street, Isfahan, Postal Code 81986, IRAN

September 19, 2007

Abstract

By exerting a magnetic field normal to a current-carrying strip we cause a deviation in the path of moving electrons as if our conducting path is a wire along an edge of the strip which at a point is deflected normal to the edge and reaches the other edge and afterwards is continued along this other edge in the same direction as before. It is clear that connecting the two end points of the transverse part of such a wire by a minor wire we expect a part of the main current to pass through this minor wire. The direction of such a current flowing in the minor wire is such that as if the current-carrying charges in the main current-carrying strip are positive charges. This is the basis of this article. A quite practical experiment is proposed for testing the presented theory.

1 Introduction

In the Hall effect it is observed that when we exert a magnetic field normal to the surface of a current-carrying conducting strip, if two opposite points of the two edges of the strip is connected together by a minor wire, an electric current will flow in this wire. Provided that the direction of current in the strip remains unchanged, it is seen that the direction of current in the minor wire depends on the kind of the strip.

What is noticeable in this effect is that existence of two different directions in the wire, which depends on the kind of the current-carrying strip, in the first instance indicates that the kind of the carriers of current in the strip depends on the kind of the strip, ie in some cases, as usual, electrons are the carriers of current, and in other cases positive charges must be carriers of current.

There is no dispute over the mechanism of carrying current by the electrons. Difficulty in explaining the mechanism of carrying current will appear if positive charges are to be carriers of current assuming that positive charges in solids are stationary. To obviate this difficulty, it is claimed that it is not in fact the positive ions that flowing in a wire cause an electric current, but it is the holes or in fact empty places of electrons that with an apparent motion carry current as if positive ions are doing so.

It is clear that when the empty places of electrons (or holes) are to be displaced, indeed the electrons move in the same their usual direction, and in fact nothing new occurs. Furthermore, while it is quite reasonable that a moving electron in a magnetic field endures a force (normal to both the electron path and the magnetic field), it is quite irrational that an empty place (of an electron or a hole), which is seemingly being displaced in a magnetic field, feels any force causing its motion towards one of the two edges of the strip, because, force can be exerted only on substance not on "nothing" (ie on empty place or hole)!

As an attempt to justify carrying of current by the displacement of holes, maybe it can be said that although it is in fact the moving electrons that in the current-carrying strip in the magnetic field are drawn towards one of the two edges of the strip, but crystalline structure of the strip and of the minor wire are such that when for continuation of the current an electron, in order to flow in the strip, leaves its atom just at the edge of the strip in the point of connection of the minor wire, its empty place has such electron affinity that before the next moving electron in the strip reaches this empty place abundant valence electrons of the minor wire are attracted by this hole. Therefore, we shall have an electron current in the transverse wire in the direction opposite to one expected. But such a justification is also very weak, because it is quite clear that the next electrons in the strip which due to the existence of current in the strip are moving towards this empty place are much closer to this empty place than the valence electrons of the minor wire. And in principle certainly the next electrons in the strip, which are themselves moving towards the empty place, will be involved in the intense electron affinity of the empty place much more than the stationary valence electrons of the minor wire. Consequently it is quite obvious that eg there will be no reason that such an unexpected current will appear in the transverse wire if this wire is of the same kind of strip; while this is not the case in practice.

This paper has been intended to solve this difficulty of justification of the Hall effect in a logical simple manner.

2 The justification

Consider a current-carrying wire which ideally has no resistance. Connect two points of this wire by a similar wire. How is the electric current divided between these two wires? For answering this question notice Fig. 1. The two above mentioned points are points a and b in this figure. Suppose that the two wires connecting these two points have the same electrical resistivity, ρ, and their cross section areas are the same, A, and the length of the wire 1 is Ll_1 and the length of the wire 2 is l_2. In a simple electrical analysis (that can be done by the reader) in which equivalent resistances are used, it can be shown easily that always we have $I_2/I_1 = l_1/l_2$ for the ratio of the current passed through the wire 2, I_2, to the current passed through the wire 1, I_1, provided that the electrical resistivity of all the wires is considered equal to zero. Therefore, if the point b is a loop of the wire 2 sliding on the length of the wire 1, bringing the point b close to the point a the current in the wire 2 (with the constant length l_2) will approach zero, and taking the point b away from the point a the current in the wire 2 will increase (approaching half of the constant current passing through the circuit).

Now suppose that instead of the wire 1 and its continuation we have a strip of the same kind of the wire 2. Furthermore, suppose that the point a is on an edge of this strip and the point b is on the other edge of the strip; see Fig. 2. It is natural that in this state too, a similar mechanism for dividing the current is expected, ie assuming that the arrow shown in the figure indicates the direction of the current of electrons, we expect, by bringing the point b, on the edge which it is located on it, close to a point just opposite to the point a, the current in the (transverse) wire to decrease approaching zero, and taking it away, this current to increase.

That the current passing through the wire increases by taking the point b away from the point opposite to the point a on the same edge on which b is located, will be certainly true even if a magnetic field is exerted normal to the surface of the strip, because although in this state the current of electrons is drawn eg towards the edge related to the point a, but since as before the electrons are moving in the same previous direction, in any case we must expect increasing of the contribution of current flowing in the (transverse) wire when taking the point b away. Therefore, if the kind of the strip is such that when the point b is just opposite to the point a we have an electron current from b to a instead of one from a to b when exerting the above mentioned magnetic field, this will be certainly the case that by displacing the point b in the direction of the current of electrons in the strip on the same edge related to b we shall observe that gradually the current of electrons from b to a (existent in the wire) will decrease and at a point will vanish and afterwards current of electrons from a to b will appear which will increase gradually by taking the point b away more and more.

But, while there is not any current in the wire when b is opposite to a and there exists no magnetic field, why does in principle there exist current in the transverse wire when b is opposite to a and a magnetic field normal to the strip is being exerted? And why can the direction of this current be changed depending on the kind of the strip while keeping the direction of current in the strip constant? The reason can be simply that by exerting a magnetic field normal to the surface of current we cause the (current-carrying) moving electrons to be drawn towards an edge of the strip and in fact with this act we create a local deviation in the path of the current, ie exerting the magnetic field, path of the moving electrons will be one shown in Fig. 3(b) not one shown in Fig. 3(a). Now let's assume ideally that by exerting the field the current-carrying conductor will be no longer straight as shown in Fig. 4(a) (which is equivalent to the same strip), but will be the deviated (major) wire shown in Fig. 4(b). In this manner the points a and b in Fig. 4(b) will be the same points a and b in Fig. 2 if the point b is located just opposite to the point a in this figure and the magnetic field normal to the strip is being exerted in this figure. By connecting the two points a and b in Fig. 4(b) by another (minor) wire we expect according to the reasoning related to Fig. 1 that the electrons flow from b to a in this minor wire. This means that we similarly should expect that when b is located opposite to a in Fig. 2 and the normal magnetic field is being exerted the electrons flow from b to a not from a to b in the minor wire in Fig. 2 (ie just the same phenomenon which seems unusual to us and for some substances such as Fe and Al occurs).

But why in many cases do we have flowing of electrons from a to b instead of b to a in the minor wire in Fig. 4(b)? Because when the electrons rushing from b to a in the major wire in Fig. 4(b) (shown by an arrow in this figure) reach the point a, they meet a right angle which force them to turn and continue their motion upwards; consequently because of their direct collision with the point a they in fact strike an impact on this point. When these points of a and b are connected to each other by another minor wire, these impacts or in fact these exertions of electronic forces can cause current of electrons from a to b in the minor wire if they are sufficiently strong, although as we said, in the usual state this current should be from b to a in the minor wire. (Obtaining of such a situation depends directly upon the electric resistances of the materials of the current-carrying strip and minor wire and also upon the configuration of the minor wire.) But if these impacts are not sufficiently effective, such a situation will not occur and direction of the current of electrons in the minor wire will be, as before, from b to a.

The above mechanism of striking of impacts and exertion of electronic forces may seem rudimentary but is not at all invalid and has been presented in this form only for simplicity at present. Indeed such a mechanism is almost the same one presented for current justification of the Hall effect which states that the electronic current is drawn towards an edge and because of the assembling of electrons at this edge a potential difference is produced which causes flowing of current in the minor wire connecting the two edges of the strip to each other.

The mechanism presented in this article can be tested by a proposed experimental way: If really electrons rushing from b to a in the minor wire in Fig. 4(b) can exert force on the electrons of the point a and cause partial flowing of electrons from a to b in the minor wire in some cases, we can expect that they do similar work in similar cases. Suppose that we have made a right angle loop as shown in Fig. 5 from a conducting wire such that in the vertex point of the right angle the two crossing parts of the wire are welded together. It is clear that assuming that the arrows show the direction of the current of electrons, we expect that we have also an anticlockwise electron current in the loop from b to a, because in the path of the main current, b has been located before a and according to the reasoning related to Fig. 1 we expect that a part of the current to flow from b to a in the loop. But when the main current is sufficiently intense and the related conditions including resistances related to the materials of the circuit are provided properly, it is possible that impacts from the electrons of the right branch moving towards the vertex point, directly exerted on the electron of the cross-section of the circuit (or in fact of the loop) in the point a, are such strong that cause a clockwise electron current in the loop from a to b which predominates the usual anticlockwise current from b to a, and consequently we observe a clockwise current in the loop from a to b. Performing such an exact experiment can test the validity of the theory presented in this paper.

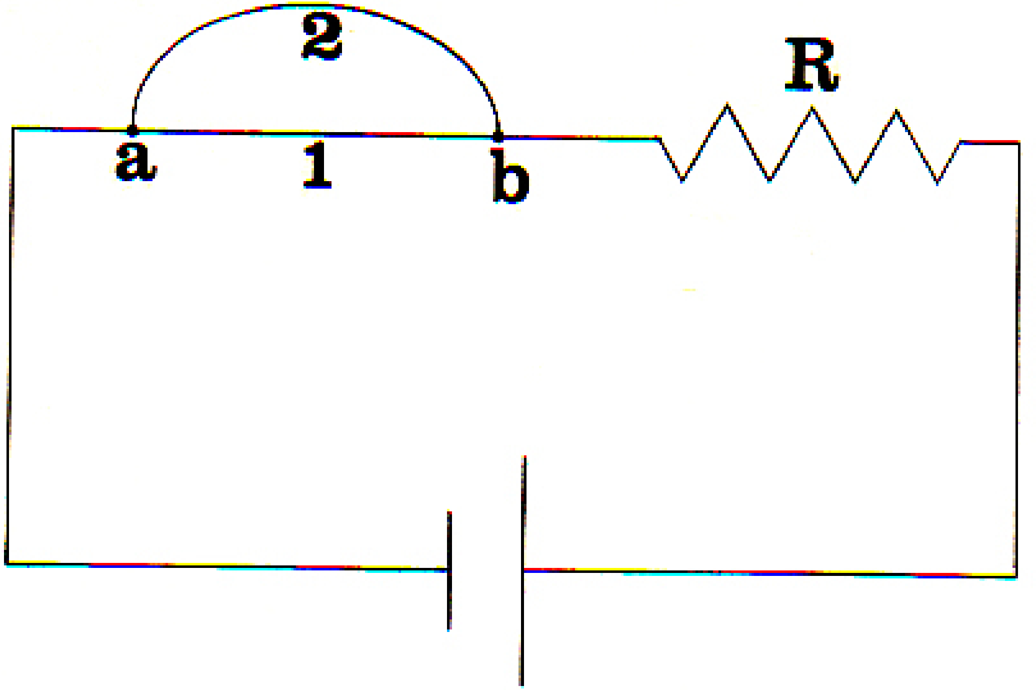

Fig. 1

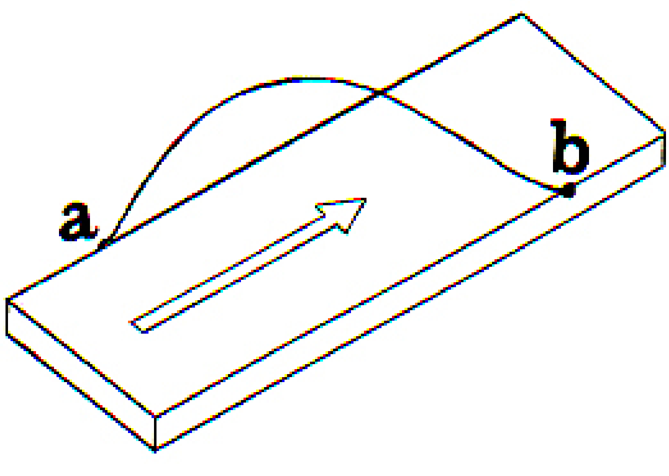

Fig. 2

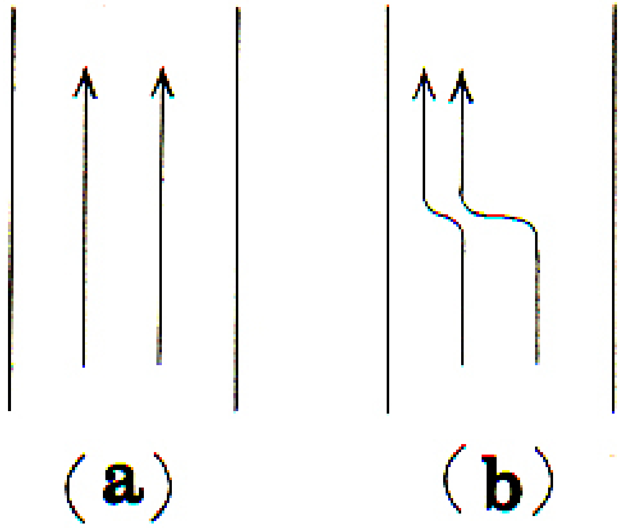

Fig. 3

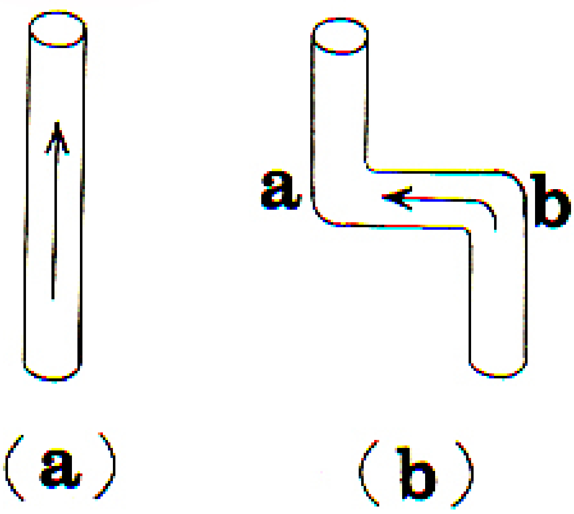

Fig. 4

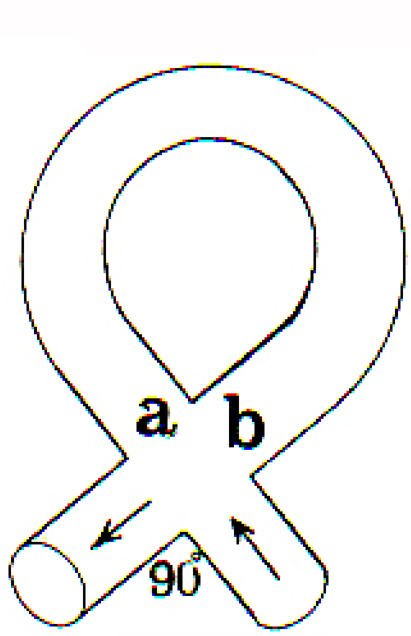

Fig. 5